Received: 20 December 2021 | Revised: 13 February 2022 | Accepted: 7 April 2022

DOI: 10.1111/zsc.12540

Zoologica Scripta | WILEY

ORIGINAL ARTICLE

Z-chromosome outliers as diagnostic markers to discriminate Mallard and Chinese Spot-billed Duck (Anatidae)

Irina V. Kulikova[1] | **Sergei V. Shedko**[1] | **Yury N. Zhuravlev**[1] | **Philip Lavretsky**[2] | **Jeffrey L. Peters**[3]

[1]Federal Scientific Center of the East Asia Terrestrial Biodiversity, Far Eastern Branch of the Russian Academy of Sciences, Vladivostok, Russia

[2]Department of Biological Sciences, University of Texas, El Paso, Texas, USA

[3]Department of Biological Sciences, Wright State University, Dayton, Ohio, USA

Correspondence
Irina V. Kulikova, Department of Evolutionary Zoology and Genetics, Federal Scientific Center of the East Asia Terrestrial Biodiversity, Far Eastern Branch of the Russian Academy of Sciences, prospekt 100-let Vladivostoku, 159, Vladivostok 690022, Russia.
Email: kulikova@biosoil.ru

Funding information
This study was funded by U.S. Civilian Research & Development Foundation (RUB1-7094-VL-13) and Far East Branch of Russian Academy of Sciences (CRDF-14-004) grants. The results were obtained with the use of IACP FEB RAS Shared Resource Center 'Far Eastern Computing Resource' equipment (https://cc.dvo.ru)

Abstract

Closely related bird species often exhibit elevated differentiation in the Z-chromosome. Genetic differentiation is repeatedly found in so-called 'islands of differentiation' that might contain loci under selection that contribute to reproductive isolation. Using double-digest restriction-associated DNA sequencing (ddRAD-seq), we examined genomic divergence of Mallard (*Anas platyrhynchos*) and Chinese Spot-billed Duck (*A. zonorhyncha*), two closely related species of ducks that occasionally hybridize. The taxa were distinguishable based on overall ddRAD-seq allele frequency differences. However, differentiation on the Z-chromosome was about 4.5 times greater than observed for autosomal DNA and included three fixed differences in SNPs. These SNPs are the first species-specific molecular markers revealed among mallard group species. The causes of elevated Z-chromosome divergence are discussed, including the possibility that Z-linked loci are resistant to introgression and potentially linked to phenotypic differences between the species.

KEYWORDS

ddRADseq, diagnostic single-nucleotide polymorphisms, Mallard, population genomics, sex chromosome, speciation

1 | INTRODUCTION

Strong phenotypic divergence in the absence of genetic divergence is not unusual among avian taxa, especially when only one or a few loci are examined (Dhami et al., 2016; Joseph et al., 2009; Toews et al., 2016, etc.). This uncoupling of phenotypic and genetic diversity suggests that selection has played a role in divergence or even speciation. Selection can likewise contribute to the uncoupling of divergence among unlinked loci, resulting in heterogeneous patterns of divergence across the genome. Indeed, discord among mitochondrial DNA (mtDNA), autosomal and sex chromosome markers is often discovered at early stages of speciation or in evolutionarily young, recently radiated bird species (Dhami et al., 2016; Ellegren et al., 2012; Lavretsky, DaCosta, et al., 2015). Furthermore,

additional evolutionary processes, including genetic drift, gene flow and mutations, contribute variably to the divergence landscape of markers with contrasting modes of inheritance and different effective population sizes (Mayr, 1982; Noor & Feder, 2006). Plumage divergence coupled with heterogeneous genomes is well illustrated by a group of recently radiated ducks, the mallard group (*Anas* spp., Lavretsky et al., 2019).

Mallard *Anas platyrhynchos* is the most numerous Holarctic game waterfowl species, and it is closely related to 13 species comprising the mallard group. Whereas the Mallard is a sexually dichromatic migratory duck, the other 13 species are mainly monochromatic, non-migratory species that have narrower local ranges and smaller population sizes. The origin, diversification and evolution of mallard group species have long been the target of molecular genetics research (Brown et al., 2019; Ford et al., 2017; Kulikova et al., 2004; Lavretsky et al., 2014, 2015; McCracken et al., 2001; Rhymer et al., 1994; Wells et al., 2019). North American mallards have been studied especially thoroughly and paradigmatically with a variety of molecular genetic markers (Lavretsky, DaCosta, et al., 2015, 2019; Lavretsky et al., 2014, 2019, 2020; McCracken et al., 2001; Peters et al., 2013, 2016). Despite being phenotypically diagnosable and more or less structured at mitochondrial DNA, species within the mallard group have little to no phylogenetic structure recovered from nuclear introns (Lavretsky, Hernández-Baños, et al., 2014; Lavretsky, McCracken, et al., 2014), which is probably due to recent speciation and gene flow. In recent years, with the development of high-throughput sequencing methods that allow the simultaneous analysis of multiple loci scattered through the genome, it has become evident that despite high genetic similarity and recurrent gene flow, the Mallard and allies possess some genomic regions that appear to be under divergent selection (Lavretsky, DaCosta, et al., 2019; Lavretsky, Janzen, et al., 2019; Lavretsky et al., 2020). These regions of high differentiation might contribute to species integrity and prevent species merging.

Genome-scale studies of the mallard group have mostly focused on the North American species which comprise a recently radiated clade of closely related species and subspecies (Lavretsky, DaCosta, et al., 2019; Lavretsky, Hernández-Baños, et al., 2014; Lavretsky, McCracken, et al., 2014). Here, we expand on this comparison by examining genomic divergence between the Mallard and a member of a presumably separate Eurasian radiation. There are three mallard group species in Eurasia: Mallard, Chinese or Eastern Spot-billed Duck *Anas zonorhyncha* (hereafter Spot-billed Duck, spotbill), and Indian Spot-billed Duck *Anas poecilorhyncha*. The Mallard has a vast breeding range across the continent from the Russian Far East through Siberia, Ural to Europe. Chinese Spot-billed Duck's breeding range is limited to East Asia; it extends from the southern Sakhalin to the west coast of eastern Siberia's Lake Baikal and includes most of Japan, Korea and north-eastern China (Figure 1). Breeding ranges of Mallard and Chinese Spot-billed Duck overlap in the south of Russian Far East and north-eastern China. Similar to the situation in North America, several Mallard x Chinese Spot-billed Duck hybrids are shot almost annually in the Primorye region (south of Russian Far East; unpublished data), which is not surprising due to their close evolutionary relationship, similar behaviour, morphology, ecology and high rate and frequency of interspecific hybridization in the Mallard complex (Kulikova et al., 2005; Kulikova & Zhuravlev, 2009).

Chinese Spot-billed Duck has a diagnosable phenotype, but it is not clearly differentiated from Mallard based on mtDNA (Kulikova et al., 2004) and is indistinguishable from Mallard at autosomal introns (Fst = 0.014; range for 13 introns = −0.02 to 0.07; Wang et al., 2019). However, some divergence between Mallard and Chinese Spot-billed Duck caused by the difference in major allele frequencies was observed at the Z-linked locus CHD1Z (Fst = 0.15; Wang et al., 2019). Whole genome sequencing revealed low genetic differentiation (Fst ~0.01), admixture and close genetic relationship between these taxa (Guo et al., 2021). Cost-effective double-digest restriction-associated-DNA sequencing (ddRAD-seq) methods proved to be extremely useful in discovering and genotyping scores of single-nucleotide polymorphisms (SNPs) randomly distributed across the genome in non-model species, identification of gene flow, hybrid individuals, founder events and singling out divergent genomic regions under selection (Lavretsky, DaCosta, et al., 2015, 2019; Lavretsky, Engilis, et al., 2015; Lavretsky, Janzen et al., 2019; Peters et al., 2016; Peterson et al., 2012).

The primary objective of this study was to use ddRAD-seq to examine genetic differentiation between Mallard and Chinese Spot-billed Duck. More specifically, we address the following questions: (a) Are there any divergent species-specific loci or just allele frequency differences sufficient to distinguish these species, (b) on which chromosomes are such loci primarily located, (c) is there any evidence of diversifying selection acting on them or is the divergence explained by other evolutionary factors in accordance with the marker's effective population sizes.

2 | MATERIALS AND METHODS

2.1 | Sampling, DNA extraction and ddRAD-seq library preparation

Liver or muscle tissues were obtained from 23 Spot-billed Ducks (17 males, six females), 29 Mallards (19 males, 10 females) and three male Mallard x spotbill hybrids (Figure

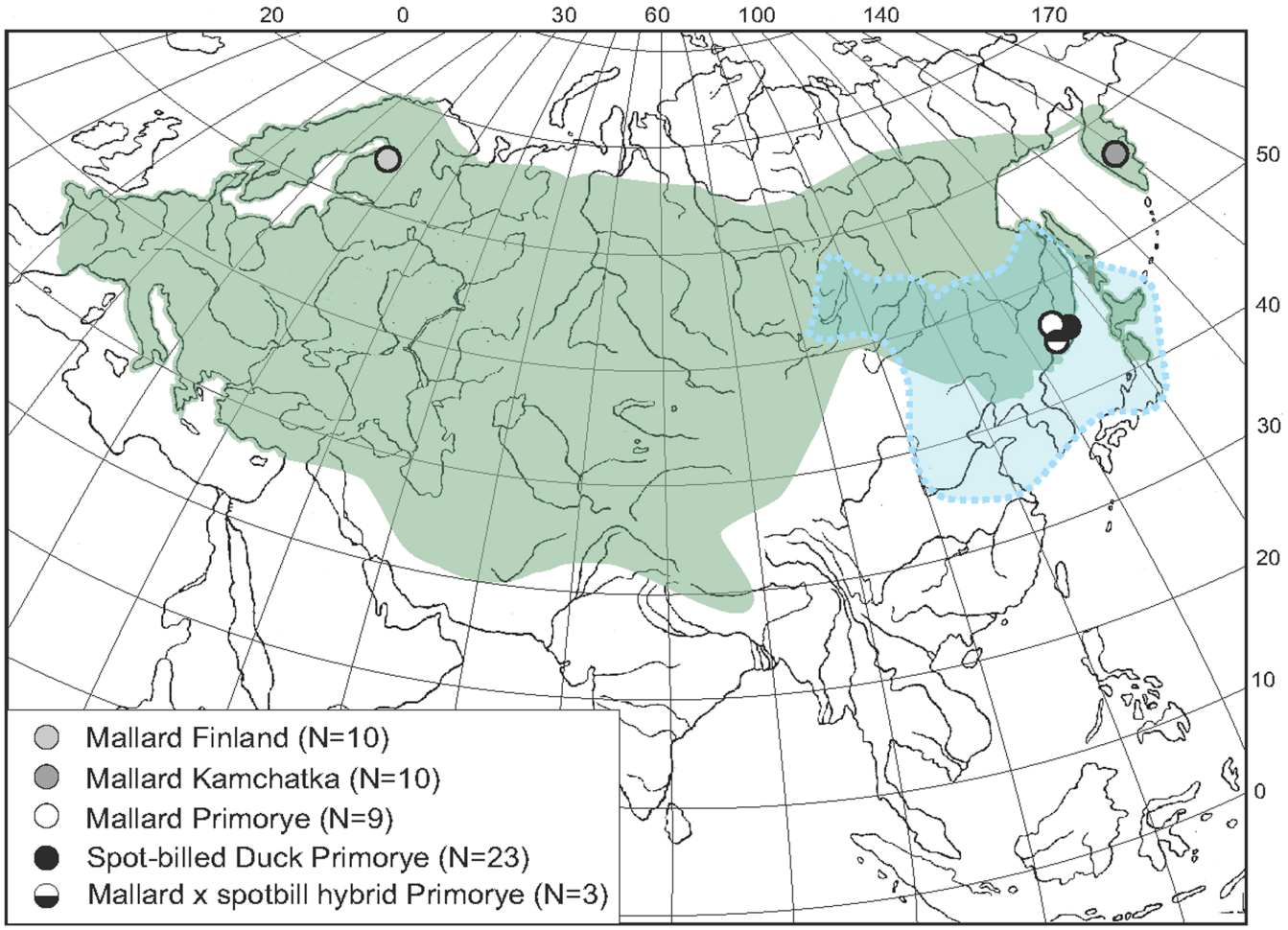

FIGURE 1 Map of sample locations for Chinese Spot-billed Ducks, Mallards and hybrids. *N* = number of samples. Species areas are colour coded: Chinese Spot-billed Duck (blue), Mallard (green). See Table S1 for additional sample information

S1, Table S1). We extracted genomic DNA using a DNeasy Blood & Tissue kit following the manufacturer's protocol (Qiagen).

ddRAD-seq libraries were prepared following the protocol of DaCosta and Sorenson (2014) to generate a pseudorandom sample of DNA sequences scattered throughout the genome. In brief, 0.5–1 μg of genomic DNA was digested using the restriction enzymes SbfI and EcoRI (10 U of each), and a unique combination of barcoded and indexed adapters was ligated to the digested DNA for each individual. We size-selected ligated DNA by excising 300–450 bp fragments from a 2% low-melt agarose gel. The size-selected DNA was purified using a MinElute Gel Extraction Kit (Qiagen) and amplified using standard PCR with Phusion high-fidelity DNA polymerase (Thermo Scientific). PCR products were purified using magnetic AMPure XP beads (Beckman Coulter, Inc.) and quantified using real-time PCR and an Illumina library quantification kit (KAPA Biosystems). Equimolar concentrations of each individual library were pooled and sequenced (150 base pair reads) on an Illumina HiSeq 2500 at TUCF Genomics, Tufts University (Medford). Raw Illumina reads have been deposited in NCBI's Sequence Read Archive (SRA; http://www.ncbi.nlm.nih.gov/sra; SRA BioProject PRJNA828171, SAMN27642144-SAMN27642198).

The raw Illumina reads were demultiplexed and processes using the DaCosta and Sorenson's (2014) computational pipeline. For each individual, identical reads were collapsed while retaining read counts and the highest quality score for each position. Individual reads with an average Phred score <20 were removed. The retained reads from all individuals were then combined and clustered into putative loci, and the highest quality read from each cluster was compared to a Mallard reference genome (accession numbers SS263068950 – SS263191362, Huang et al., 2013; Kraus et al., 2011). Clusters with the same BLAST hits were combined, and the reads for each cluster were aligned using MUSCLE v. 3 (Edgar, 2004). Alignments were manually checked and edited as described in Peters et al. (2016).

The final alignments were used to genotype individuals at each locus using the following criteria from DaCosta

and Sorenson (2014). Specifically, individuals were scored as homozygous at a locus if ≥93% of reads were consistent with a single haplotype, and heterozygous if ≥29% were consistent with a second haplotype. Individuals were also scored as heterozygous if as few as 20% of reads were consistent with a second allele and that haplotype was represented in other individuals. Individual genotypes were flagged if they did not meet either of these criteria, had evidence of more than two haplotypes or were represented by ≤5 reads; for those genotypes, we retained only the allele represented by the majority of reads and scored the second allele as missing data. We retained for analysis all loci with ≤10% missing genotypes and ≤5% flagged genotypes.

A representative sequence from each of the final alignments was aligned to two reference Mallard genomes (assembly version IASCAAS_PekingDuck_PBH1.5, accession number GCF_003850225.1; assembly version CAU_duck1.0, accession number GCA_002743455.1), which allowed us to localize each locus and separate autosomal and Z-linked loci for downstream analyses. Only uniquely aligned reads were selected.

2.2 | Population structure

We used the program STRUCTURE 2.3.4 (Falush et al., 2003) to estimate the number of genetic populations (*K*) and to assign individuals to those populations. STRUCTURE, based on 3130 autosomal and 194 Z-chromosome loci, was run using an admixture model and correlated allele frequencies among populations without prior information regarding sampling locations. We ran 20 replicates for each value of *K* in the range of 1–4 with 200,000 steps of the Markov chain Monte Carlo (MCMC) after a burn-in period of 50,000 iterations. To display the results in a graphical interface, we processed STRUCTURE output files in CLUMPAK v.1.1 (Kopelman et al., 2015). The optimum *K* was determined using Δ*K* (Evanno et al., 2005) calculated in the CLUMPAK v.1.1. To further assess population structure, we used principal coordinates analysis (PCoA) based on the Euclidian distances between individual genotypes and implemented by *dudi.pca* in the R software package *Adegenet* v.2.1.3 (Jombart, 2008). The dissimilarity matrix calculation was premised on loci variability. PCoA plots were produced using the *ggplot2* package version 3.3.2 (Wickham, 2016). Then, we obtained maximum likelihood estimates of population assignments for each individual with ADMIXTURE v.1.3.0 (Alexander et al., 2009). Data formatting for ADMIXTURE was done with PLINK (Purcell et al., 2007). We analysed 9139 autosomal and 220 Z-chromosome biallelic SNPs separately with 20-fold cross validation in each ADMIXTURE analysis. Because ADMIXTURE does not accommodate haplodiploid data, females were treated as being homozygous for Z-chromosome loci. The optimum *K* was based on the lowest average of CV errors across all the analyses per *K*. To process and visualize ADMIXTURE outputs, we used CLUMPAK v.1.1. ADMIXTURE plots were drawn with R Graphics package v. 3.6.3.

2.3 | Genomic differentiation and outlier analysis

To assess the pattern of genetic differentiation across the genome, we calculated the composite and pairwise per locus values of Φst, absolute divergence (i.e. d_{XY}; Nei & Li, 1979), nucleotide diversity and Tajima's D in the r package *PopGenome* (Pfeifer et al., 2014). We also calculated the ratio of adjusted Z diversity (Z diversity divided by an estimate of 1.1 for substitution rate ratio of Z vs. autosomes) to autosome diversity as an estimate of the ratio of effective population sizes of the two chromosome types (Irwin, 2018). We plotted Z-linked pairwise Φst values per locus by chromosomal position and both *d*xy and Tajima's *D* values against nucleotide diversity in Excel. Violin plots of nucleotide diversity and absolute divergence were constructed using PAST statistical software v. 4.05 (Hammer et al., 2001). BayeScan v. 2.1 (Foll & Gaggiotti, 2008) was run to identify outlier loci. The BayeScan analysis was carried out for 20 pilot runs of 5000 steps each, followed by 100,000 burn-in and 200,000 sampling steps with a thinning interval of 10. The false discovery rate (qval) was set to 0.05 and the outlier loci were plotted using the R function provided by the software. We ran two separate analyses, one with autosomal loci and another with Z-linked loci. To further estimate relationships between samples at the outlier loci, we constructed haplotype networks using the median-joining algorithm in the program Network v. 10.2.0.0 (Bandelt et al., 1999). We used a parametric *t*-test or non-parametric Kolmogorov–Smirnov (KS) test or Mann–Whitney–Wilcoxon (MWW) test to compare absolute divergence, nucleotide diversity and Tajima's *D* values between outlier and non-outlier loci identified for Chinese spot-billed ducks and mallards. Autosomal and Z-linked loci were treated in separate analyses. The *p*-value <.05 was considered statistically significant.

3 | RESULTS

After quality filtering, we recovered 3324 ddRAD-seq loci that met our coverage and missing data criteria. Among those loci, 3130 loci (397,408 aligned base pairs; 10,968 SNPs) and 194 loci (24,098 aligned base pairs; 392 SNPs)

were assigned to autosomes and the Z-sex chromosome respectively (Table S2).

Mallards and spotbills had similar overall autosomal nucleotide diversity (Table 1). Mallards Z chromosome nucleotide diversity was slightly higher than that of spotbills. Composite absolute divergence *d*xy was almost equal or slightly higher than nucleotide diversity for both types of markers. Genetic differentiation between species was higher for the Z chromosome than for the autosomes with the observed Φ_{stZ} to Φ_{stA} ratio of 4.5 (Table 1). The frequency distribution of pairwise Φ_{st} values revealed a long tail that extended into high Φ_{st} values (>0.3) for the Z-chromosome (Figure 2). The estimated $N_{e,Z}/N_{e,A}$ ratios were 0.4012 and 0.3191 for Mallard and spotbill respectively.

3.1 | Population structure

The first coordinate axis in PCoA separated Mallards and spotbills most effectively when both autosomal and Z chromosome markers were analysed (Figure 3a). The hybrids occupied intermediate positions between Mallard and spotbill samples. Using only autosomal loci resulted in lower separation among taxa; however, Mallards and spotbills still clustered into separate groups (Figure 3b). In both analyses, there were three pairs of samples (one pair of Mallards from Finland, one pair of Mallards from Kamchatka and one pair of spot-bills) that each clustered away from the main clusters. These pairs were likely close kin (Peters et al., 2016). PCoA based on Z loci distance matrix differentiated Mallards and spotbills, but the samples were widely scattered in two dimensions (Figure 3c). There was no differentiation among groups of Mallards: Finland, Kamchatka and Primorye samples were intermixed (Figure 3a–c).

Admixture analysis was based on 9139 autosomal and 220 Z-chromosome biallelic SNPs, excluding singletons. For autosomal loci, the optimum value of K was 1 with the cross-validation value (CV) of $K = 2$ being only slightly higher than the CV of $K = 1$ (Figure S1C). Admixture of $K = 2$ for autosomal markers did not discriminate clearly between samples from Mallard and Spot-billed Duck, although samples of each species consistently had higher assignments to their respective population. Hybrids could be equally identified as Mallards or spotbills (Figure S1A). The optimum value of K for Z-linked loci was 2 (Figure S1C) at which Mallard and Spot-billed Duck samples were unequivocally assigned to different populations with hybrids having intermediate ancestry in comparison with parent species (Figure S1B). Using autosomal loci admixture identified 16 admixed Mallards and 10 admixed spotbills with Q values of 22–51% and 24–35%, respectively, while analysis based on Z-linked markers ascertained mixed ancestry in eight Mallards and one spotbill with Q values of 4%–29% and 2.7% respectively. Analysis of both types of markers did not reveal any genetic structure in Mallards (Figure S1A,B).

Structure corroborated admixture results. Delta K calculation indicated that $K = 2$ was the best supported model for both autosomal and Z-linked loci. The two-population model revealed clear separation between Mallards and spotbills that received high assignment probabilities for their respective groups (Mallard: $Q > 0.997$ for autosomal loci, $Q > 0.935$ for Z-linked loci; Spot-billed Duck: $Q > 0.763$ for autosomal loci, $Q > 0.981$ for Z-linked loci; Figure 4). Hybrids showed evidence of admixture from Mallards and spotbills (assignment probabilities for autosomal loci: hybrids ← Mallards $Q = 0.840$, range 0.664–0.998, hybrids ← spotbills $Q = 0.160$, range 0.002–0.336; for Z-linked loci: hybrids ← Mallards $Q = 0.445$, range 0.280–0.616, hybrids ← spotbills $Q = 0.555$, range 0.384–0.720).

3.2 | Genomic differentiation and outlier loci

Comparing allele frequencies between Mallard and Spot-billed Duck populations, BayeScan identified autosomal and Z-linked loci under diversifying and balancing or purifying selection (Table S3). The Z-chromosome contained approximately a 2.5-time higher percentage of loci under diversifying selection than autosomes (1.5% and 0.6% respectively). It harboured three significant outlier loci with prominent Φst estimates (>0.89) and signatures of divergent selection ($\alpha > 1.26$, $q < 0.009$): (~31,239,684; ~35,972,534; ~78,252,210 base positions; Figure 5) and 12 loci under balancing or purifying selection. Five autosomal chromosomes contained nine outlier loci with high Φst (>0.26) and signatures of divergent selection (Chromosome 1 [~71,072,809 base position, ~83,509,713 base position, 2Mb region between

TABLE 1 Genetic diversity and differentiation of Mallard and Chinese spot-billed Duck. π-nucleotide diversity, *d*xy—absolute divergence, A—autosomal, Z—Z-chromosome

	π_A	dxy_A	π_Z	dxy_Z	Φ_{st}	Φ_{stA}	Φ_{stZ}
Mallard	0.00673	0.00690	0.00297	0.00295	0.02197	0.0208	0.0937
spotbill	0.00678		0.00238				

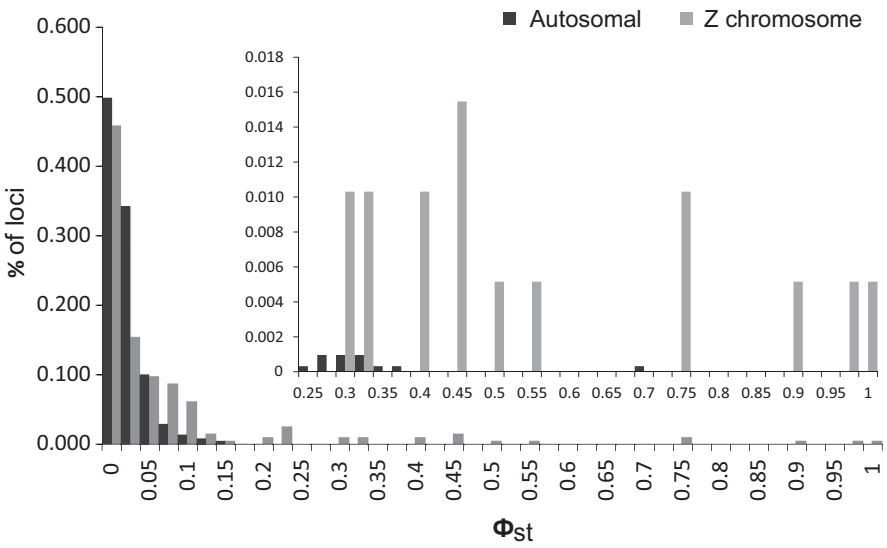

FIGURE 2 Frequency distribution of Φ_{ST} estimates across 3130 autosomal and 194 Z-linked ddRAD-seq loci between Mallard and Chinese Spot-billed Duck. Insert provides expanded view of the frequency distribution of Φ_{ST} estimates >0.25

base pair positions 1.15E8–1.17E8], Chromosome 2 [~901,937 base position], Chromosome 4 [~38,654,056 base position], Chromosome 5 [~43,757,329 base position] and Chromosome 12 [~10,622,270 base position]). Outliers with lower Φst (0.02–0.23) and evidence of divergent selection were found on chromosomes 1, 2, 4, 10, 14, 15, 21 and 29 (Table S3).

Haplotype networks for the three prominent Z-linked outlier loci demonstrated fixed allele differences and presence of species diagnostic alleles (positions 31,239,684; 35,972,534), or significant allele frequency differences (position 78,252,210; Figure 6). Locus Z:31,239,684 included a fixed difference between spotbills (100% of spotbills shared the same allele) and Mallards (96.5% of Mallards shared one allele; 3.5% had another); hybrids carried one copy of each of the two common alleles. Similarly, all the spotbills (100%) had one allele of Z:35,972,534, while 98% of all the Mallards had another allele, and there was a fixed difference between species. Hybrids again had both types of Z:35,972,534 alleles in equal proportions (50:50). Two prevalent alleles of Z:78,252,210 had considerable species-specific composition with one comprising 91.5% of all Mallards, 2.6% of spotbills and 50% of hybrids and the other one consisting of 97.4% of all spotbills, 2.1% of Mallards and 50% of hybrids (Figure 6). Φst values corresponded to network patterns; extremely high Φst (0.97, 0.98 and 88.7) were observed for all pairwise comparisons between Mallards and spotbills for loci Z:31,239,684; Z:35,972,534; and Z:78,252,210 respectively.

Z-chromosome significant outlier loci were checked against full genome sequencing reads of Mallards and spotbills available in NCBI Sequence Read Archive as possible species-specific markers (Table S4). SNPs in position 130 of ddRAD-seq locus Z:35,972,534 and in position 106 of locus Z:78,252,210 effectively differentiated Mallards and spotbills. All the Mallard and Spot-billed Duck genomes analysed had alternative nucleotide variants in those positions. Using SNP Z:31,239,684 (42A/G) as a diagnostic marker was complicated due to it having high similarity to sequences on autosomes, which had an A in position 42, in both species. Nevertheless, all matching Mallard reads had an A in position 42 of Z:31,239,684, whereas spotbill reads had a two- to four-time higher frequency of G than A in this position (Table S4). DNA reads with 42A found in Spot-billed Duck genomes were shorter than the ddRAD-seq locus Z:31,239,684 and sometimes had insertions. No such ambiguity occurred with ddRAD-seq markers.

Scatter plots allowed us to visually compare the results of calculating Tajima's *D*, nucleotide diversity and absolute divergence for outlier and non-outlier markers (Figures S2 and S3). The distribution of Tajima's *D* was similar between autosomal outlier and non-outlier markers for Mallard (KS test $p = .058$) but was not so for Spot-billed Duck (KS test $p = .031$). Tajima's *D* distributions of Z-linked outliers and non-outliers were quite similar (MWW test $p > .61$) and were shifted towards negative values in both species (Figure S2). Nucleotide diversity estimates of Z-linked outlier markers were three to four times lower on average than those of non-outlier markers in both Mallard and Spot-billed Duck (MWW test $p < .05$). In contrast, autosomal outlier and non-outlier markers had similar nucleotide diversity estimates in both species (two-tailed *t*-test $p > .7$; Figure S3). Values of *d*xy were relatively low and did not differ between autosomal outliers and non-outliers (two-tailed *t*-test $p = .102$, Figure S3), whereas average *d*xy was about three times higher for Z-linked outlier loci compared to Z-linked non-outliers (MWW test $p = .019$). Overall, compared to non-outliers, the outlier regions on the Z-chromosome were characterized by high absolute divergence and low nucleotide

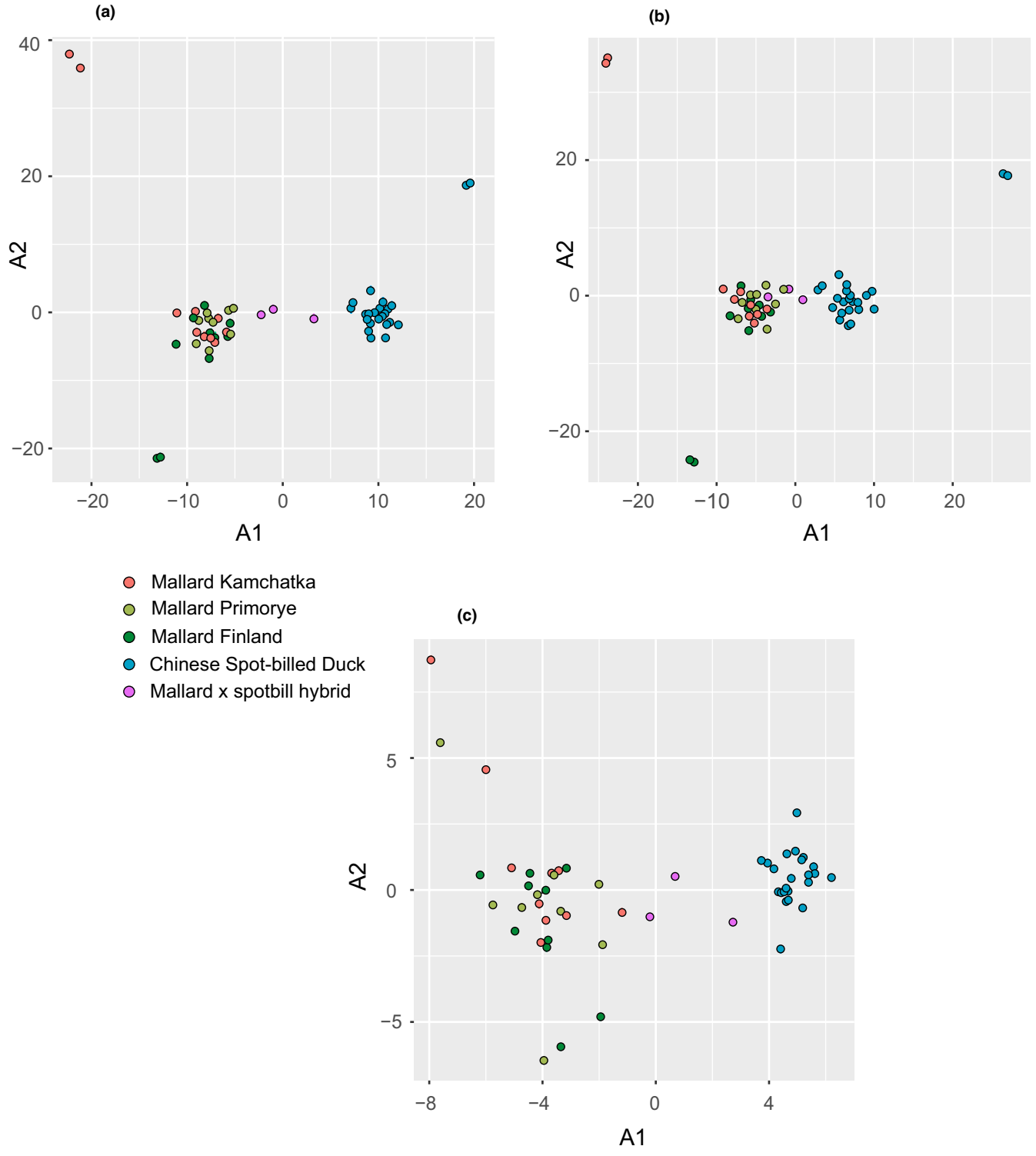

FIGURE 3 Scatter plots of the first two principal coordinates for (a) 3130 autosomal and 194 Z-linked (b) 3130 autosomal (c) 194 Z-linked ddRAD-seq loci for Mallard, Chinese Spot-billed Duck and their hybrids

diversity. Autosomal outliers were similar to autosomal non-outlier regions for these parameters.

There was a strong positive correlation between autosomal absolute divergence and nucleotide diversity in both spotbills and Mallards, and in nucleotide diversity between species (Figure S3). The correlation between these parameters for Z-linked loci was weaker, especially when comparing nucleotide diversity between spotbills and Mallards. To compare the distributions of absolute divergence and nucleotide diversity in Mallard and Spot-billed Duck, we constructed violin plots (Figure 7). The distributions of nucleotide diversity and absolute divergence

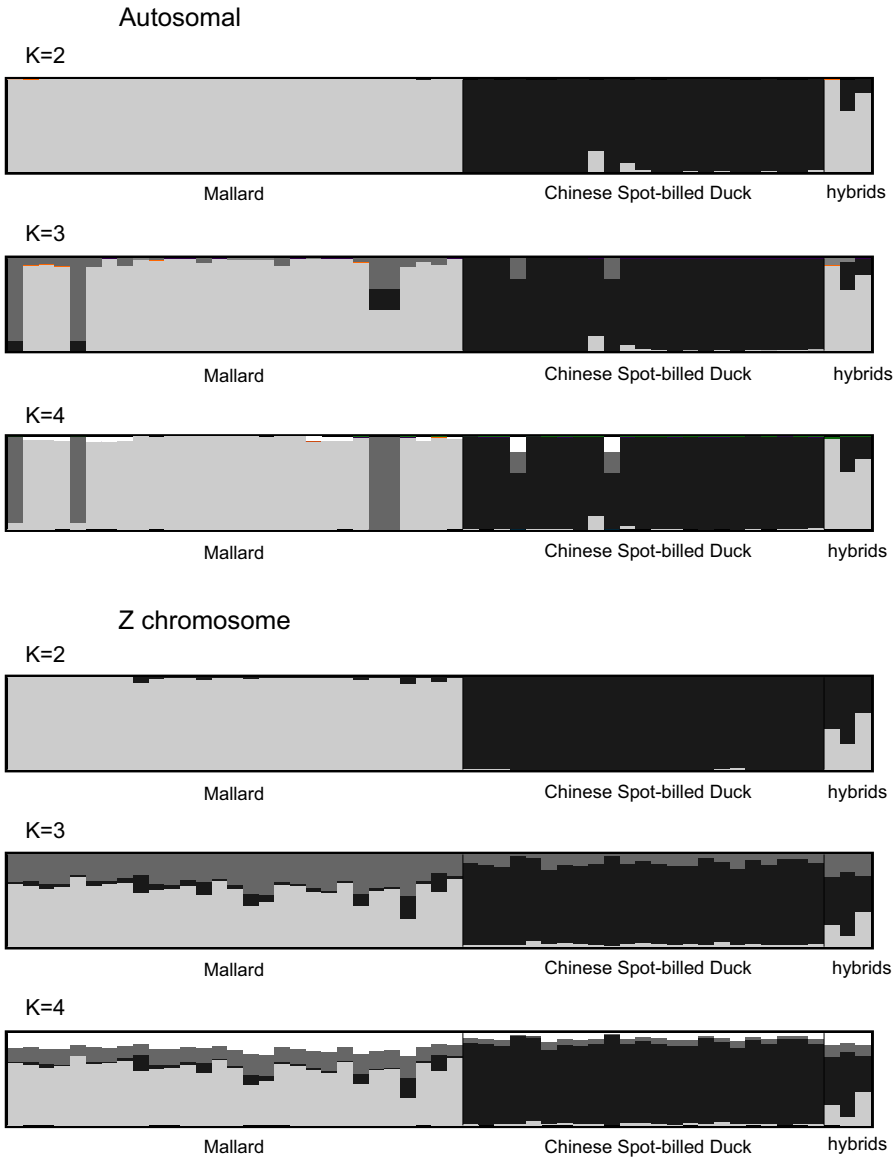

FIGURE 4 STRUCTURE assignment probabilities for mallard, Chinese Spot-billed Duck and Mallard x Chinese Spot-billed Duck hybrids for 3130 autosomal and 194 Z-linked ddRAD-seq loci. *K*—number of populations

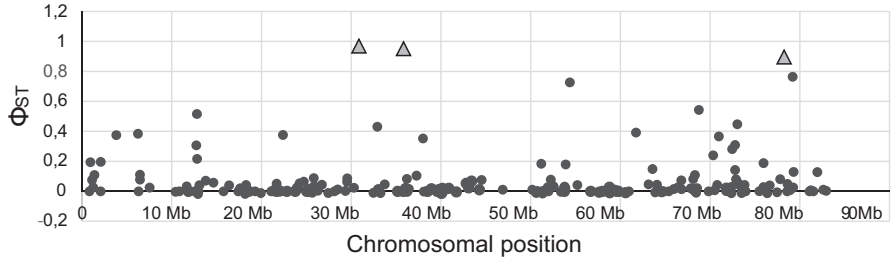

FIGURE 5 Frequency distribution of Φ_{ST} estimates for Z-chromosome with significant outliers (triangles) for pairwise comparisons between Mallards and Chinese Spot-billed Ducks

were similar for autosomal loci (ANOVA; $p = .83$) but differed significantly for Z-linked loci (ANOVA; $p = .047$). Similarly, the distributions of nucleotide diversity for Z-linked loci differed significantly between Mallard and spotbill (MWW test $p = 1.79E\text{-}05$; KS test $p = 3.96E\text{-}07$). The same was true for comparison of Z-chromosome nucleotide diversity and absolute divergence for spotbills (MWW test $p = 2.94E\text{-}06$; KS test $p = 6.80E\text{-}08$) but not for Mallards (MWW test $p = .96$; KS test $p = .76$).

4 | DISCUSSION

4.1 | Population structure

Mallard and Chinese Spot-billed Duck are waterfowl species with overlapping breeding areas in East Asia. Their hybrids are not often but quite regularly found in the south of the Russian Far East during the breeding season since the north-western expansion of spotbill's breeding

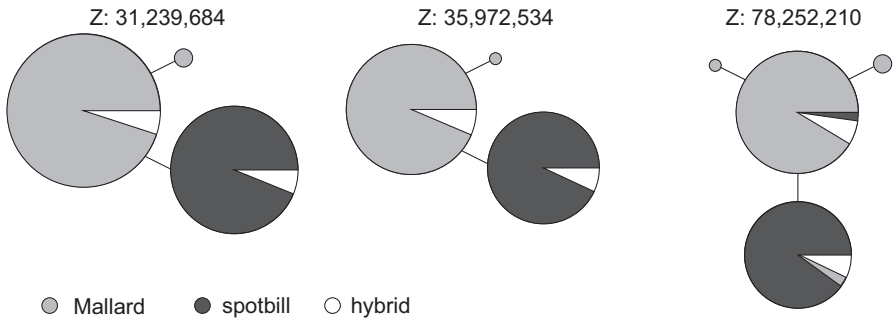

FIGURE 6 Haplotype networks for the three most extreme outlier loci on Z-chromosome (positions 29,601,875; 35,972,534; 78,252,210). Each haplotype network includes one (females) or two (males) alleles per individual

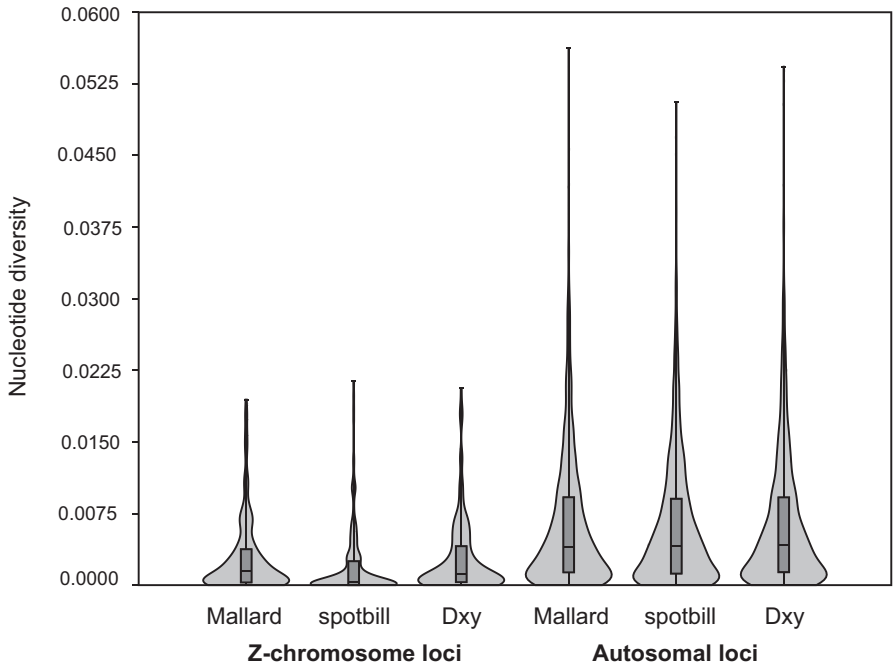

FIGURE 7 Violin plots showing the distribution of the nucleotide diversity and pairwise absolute divergence values of Z-chromosome and autosomal ddRAD loci in Mallard and Chinese Spot-billed Duck

range into Eastern Siberia that began about 80 years ago (Polivanova, 1971; Skryabin, 1963). Nevertheless, there is not any evidence of increasing numbers of hybrids or introgressive hybridization threatening the species integrity of Mallard or spotbill. Previous studies revealed the close evolutionary relationships between these species and the absence of diagnostic genetic markers. Both mtDNA and nuclear introns failed to differentiate Mallards and spotbills (Kulikova et al., 2004; Wang et al., 2019). Whole-genome sequencing supported low genetic differentiation, monophyly and genomic intermixture of these taxa (Guo et al., 2021).

In this study, a sufficient number of ddRAD-seq loci with different allele frequencies was recovered which allowed multi-locus assignment of individuals to their parental species. Admixture and structure were consistent with assigning individuals to the two separate species, although structure found much less evidence of admixture (Figure 4 and Figure S1A). Z-chromosome results of admixture and structure analysis were more similar and unequivocal in assigning Mallards and spotbills to their corresponding populations (Figure 4 and Figure S1B). Admixture and structure results are also corroborated by a much higher Φst value for Z-chromosome markers (0.094) compared to autosomal loci (0.021). More pronounced differences between species based on Z-chromosome markers have been observed between many closely related bird species (Irwin, 2018), including Mallard and its North American monochromatic relatives like Mexican Duck (*Anas diazi*), American Black Duck (*A. rubripes*) and Mottled Duck (*A. fulvigula*) (Lavretsky, DaCosta, et al., 2015, 2019; Lavretsky, Engilis, et al., 2015). Elevated differentiation on the Z chromosome is usually explained by faster Z evolution, reduced effective population size, lower recombination rate and large Z effect (Irwin, 2018; Lima, 2014; Mank et al., 2010; Wright et al., 2015). Given that species in the mallard complex diverged recently (Lavretsky, DaCosta, et al., 2019; Lavretsky, Hernández-Baños, et al., 2014; Lavretsky, McCracken, et al., 2014), higher admixture for autosomal markers was likely the result of a combination of shared ancestral variation due to the larger effective population size and perhaps recurrent gene flow.

Three phenotypic hybrids had ~50/50 mixed ancestry based on admixture analysis of Z-chromosome loci. For autosomal loci assignment of hybrids to parental species (Mallard/Spot-billed Duck) varied on average from 42/58 in admixture to 84/16 in structure (Figure 4 and Figure S1). Given the strongly defined transitional plumage and bill colour as well as close to intermediate values of assignment probabilities to parental species, we conclude that the hybrids studied are likely recent generation hybrids (F1/F2 generations). Our conclusion is consistent with the decay in differentiation for outlier loci with increased backcrossing in American black duck x Mallard hybrids (Lavretsky, Janzen, et al., 2019).

We did not find evidence of population structure in Mallard, despite sampling from widespread locations across Eurasia. There was no difference observed between Finland, Kamchatka and Primorye samples which confirmed the widely accepted view on Mallard in Eurasia as one panmictic population (Kraus et al., 2013; Kulikova et al., 2012) with the exception of some population structure revealed in Europe as a result of release of game-farm Mallards in a few European countries (Söderquist et al., 2017).

4.2 | Genomic differentiation

Mallard and Spot-billed Duck had similar overall nucleotide diversity with absolute divergence being almost equal or slightly higher than within population diversity for both types of markers. The recent divergence between Mallard and Spot-billed Duck likely explains these similarities, which is in concordance with the results obtained on these taxa using mtDNA (Kulikova et al., 2004), introns (Lavretsky, McCracken, et al., 2014; Wang et al., 2019) and whole-genome sequencing (Guo et al., 2021). The dates of divergence based on different markers vary but all studies converge on the idea of a recent radiation of mallard group species in general and recent divergence of Mallard and Spot-billed Duck in particular (Guo et al., 2021; Kulikova et al., 2004; Lavretsky, McCracken, et al., 2014; Wang et al., 2019).

Within-population nucleotide diversity was depressed on the Z-chromosome of both species with the ratio of Z to autosomal nucleotide diversity being much lower than expected from population size effects. Thus, estimated $N_{e,Z}/N_{e,A}$ ratios for Mallard and Spot-billed Duck (0.4012 and 0.3191 respectively) were considerably below that of 0.75 characteristic of idealized populations and well below the theoretical minimum of 9/16 (i.e. 0.5625) expected in case of extreme variance in male mating success (Charlesworth, 2001). However, ducks are not polygynous; their prevalent mating system is seasonal monogamy. Forced copulations as a secondary mating tactic of monogamous males despite being quite common in ducks is not always successful due to morphological complexity in female genital morphology (Brennan et al., 2007) and mate defence by paired males (McKinney et al., 1983). Analysis of extra-pair paternity in Mallard revealed that extra-pair offspring accounted for 9.3% of all offspring (Kreisinger et al., 2010), which is not enough to shift the N_e of Z chromosome to achieve the observed $N_{e,Z}/N_{e,A}$ ratios.

In addition to lowered Z chromosome nucleotide diversity and $N_{e,Z}/N_{e,A}$ ratio in both species, relative divergence was 4.5 times higher for Z-linked than for autosomal ddRAD markers. We believe that linked selection would fit quite well as an explanation of the observed high $\Phi st_{Z:A}$ and low $N_{e,Z}/N_{e,A}$ ratios. Most loci under balancing or purifying selection as well as three prominent loci under positive selection were found on the Z-chromosome (Table S3). It is challenging to identify if selection was promoting lineage-specific adaptation or reproductive isolation. Given that hybridization between Mallard and Spot-billed Duck occurs on a regular basis, we assume that the pronounced differentiation of the Z-chromosome could arise from limited gene flow due to genetic incompatibilities between mutations accumulated at Z-linked loci after recent species divergence. Many sexually selected traits such as male plumage and female preference are Z-linked in birds (Irwin, 2018). Therefore, sexual selection is an equally probable explanation of the observed patterns. Mallard is a dimorphic species whereas Spot-billed Duck males and females have similar plumage colour with female Mallards. Reduction of Z-linked loci gene flow may be due to female-biased gene flow because hybrid males fail to compete for mates similar to two hybridizing Darwin's finches (Lamichhaney et al., 2020).

Spotbills had slightly lower overall Z-linked nucleotide diversity and lower $N_{e,Z}/N_{e,A}$ ratio than Mallards. Lower $N_{e,Z}/N_{e,A}$ ratio revealed in monomorphic Spot-billed Duck as compared to dimorphic Mallard is consistent with the results obtained on eight species pairs of birds with contrasting levels of dichromatism (Huang & Rabosky, 2015), contradicting the expectation of stronger sexual selection on Z-chromosome in dimorphic species (Price, 2019). Mallard and Spot-billed Duck had statistically significant differences between distributions of Z-linked nucleotide diversity. An extremely low median value and an excess of loci with low-nucleotide diversity were characteristic of spotbills (Figure 7). The reduction in nucleotide diversity was uniformly distributed along the Z chromosome (Figure S4). We hypothesize that this steady reduction of nucleotide diversity is the consequence of a bottleneck followed by strong genetic drift experienced by spotbills during the recent past. Similar to the Z-chromosome, the effective population

size of Spot-billed Duck based on mtDNA was about five times lower than that of Mallard (Wang et al., 2019). Mitochondrial DNA is haploid and has one-fourth the effective population size of autosomal DNA. The somewhat less prominent reduction in nucleotide diversity on Spot-billed Duck Z chromosome could be explained by the fact that the effective population size of Z chromosome is three-fourths that of autosomes. Whole-genome sequencing results are consistent with our assumption. PSMC analysis of Mallard and Spot-billed Duck full genomes showed that N_e of both species increased about 1 mya, reached a maximum about 40 kya and then declined dramatically near the start of the last ice age. Overall N_e in Mallard was slightly higher than in spotbills before the decline in population size and almost two times higher after the decline (Guo et al., 2021).

4.3 | Outlier loci

Three Z-linked outlier loci were strong candidates for being under divergent selection (Table S3). These outlier regions were characterized by high relative divergence, negative Tajima's *D* values, relatively high estimates of absolute divergence and low nucleotide diversity. Unlike Z-linked outlier regions, autosomal outliers and non-outliers had similar estimates of the above-mentioned parameters, and thus, did not carry the same signatures of selection as observed for Z-linked outliers (Figures S2 and S3). Therefore, we hypothesize that advanced divergence between parts of Z-chromosomes of Mallards and Spot-billed Ducks is likely to be the result of divergent or diversifying selection.

Phenotypic differences between Mallard and Spot-billed Duck involve mainly coloration traits. Given that many genes regulating plumage and bill colour are found on the avian Z-chromosome (Irwin, 2018), and that we found weak autosomal genetic differences, we hypothesize that phenotypic differences between Mallard and Spot-billed Duck are likely Z-linked. Our findings and inferences on the Z-chromosome outlier loci are consistent with the results obtained on Mallard and Mexican duck, Australian teals (Dhami et al., 2016; Lavretsky, DaCosta, et al., 2015) and other avian species. Z-chromosome islands of differentiation contain genes regulating plumage pattern and colour in parulid warblers, capuchino seedeaters, Gouldian finch (*Erythrura gouldiae*), Reunion grey white-eye (*Zosterops borbonicus*) (Bourgeois et al., 2020; Campagna et al., 2017; Kim et al., 2019; Toews et al., 2016; Toomey et al., 2018). However, limited genomic coverage provided by ddRAD sequencing (~0.03%–0.04% of the genome) could prevent the discovery of pronounced islands of differentiation in our data.

Lavretsky, DaCosta, et al. (2019) detected evidence of an island of divergence on the Z-chromosome that was located around positions 1.7E7–3.8E7 bp. This region of elevated divergence was observed within all pairwise comparisons between Mallard and its North American relatives. In contrast, for Mallard versus spotbill, loci with moderate to high Φst values were scattered across the Z-chromosome. Notably, however, two of the significant outlier loci fell within the same region of elevated divergence detected by Lavretsky, DaCosta, et al. (2019) on the Z-chromosome. Thus, this is a good candidate region for discovering loci that are under divergent selection and perhaps involved with speciation or phenotypic differences.

This is a first time that species-specific SNPs have been discovered between any species within the mallard group. Three diagnostic SNPs found in Z-linked outlier regions could be effectively used as species-specific molecular markers in research with these species. In order to check their applicability in species assignment, we ran Megablast search across fully sequenced Mallard and Spot-billed Duck genomes deposited in GenBank (Table S4). The results of Megablast strongly supported the idea of using these SNPs as species-specific markers except for the confusion with the SNP (42A/G) in outlier region Z:31,239,684. All Mallard reads had an A in position 42, whereas spotbill reads had G and less frequently A. The latter was due to the presence of similar sequences (with identity <85%) on chromosomes 1, 2, 3, 5, 9, 12, 14 etc. in both species. Autosomal sequences identical to Z:31,239,684 with an A in position 42 were shorter than the outlier region and therefore did not obstruct spot-billed duck identification based on ddRAD-sequencing. The other SNPs worked perfectly well-discriminating Mallards and spotbills. Three Mallard x Spot-billed Duck hybrids were heterozygotes at three diagnostic SNPs, supporting our assumption on their recent hybrid origin (i.e. they were likely F1/F2 generation hybrids).

5 | CONCLUSION

In conclusion, elevated differentiation on the Z-chromosome between Mallard and Spot-billed Duck is likely the result of interactions among multiple evolutionary mechanisms and factors such as selection, demographic history, genetic drift, gene flow, mating system, sex-biased dispersal etc. We suggest that the most influential among them were genetic drift and selection. Z-linked diagnostic SNPs appear to be associated with regions of the Z chromosome under divergent selection. We hypothesize that such regions play an important role in resistance to introgression and thus prevent gene flow between

recently diverged Mallards and spotbills. Whether these regions are involved in phenotypic differences between the species and sexual dimorphism is the prospect of future work. We believe that whole-genome sequencing along with plumage analyses will shed light on phenotypic evolution and help to identify speciation mechanisms in Mallard and Chinese Spot-billed Duck.

ACKNOWLEDGMENTS

None.

AUTHOR CONTRIBUTION

JLP, IVK and YNZ conceptualized and collected data. SVS, JLP, PL and IVK analysed and supported data acquisition. IVK, SVS, YNZ and JLP equally contributed to the writing of this manuscript.

DATA AVAILABILITY STATEMENT

Raw Illumina reads have been deposited in NCBI's Sequence Read Archive (SRA) under BioProject PRJNA828171. Other data files (FASTA files; BAYESCAN, ADMIXTURE, and STRUCTURE input files) have been made available on FigShare, at: https://doi.org/10.6084/m9.figshare.19612893

ORCID

Irina V. Kulikova https://orcid.org/0000-0003-4847-2560

REFERENCES

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*, 1655–1664. https://doi.org/10.1101/gr.094052.109

Bandelt, H. J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, *16*, 37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036

Bourgeois, Y. X. C., Bertrand, J. A. M., Delahaie, B., Holota, H., Thébaud, C., & Milá, B. (2020). Differential divergence in autosomes and sex chromosomes is associated with intra-island diversification at a very small spatial scale in a songbird lineage. *Molecular Ecology*, *29*, 1137–1153. https://doi.org/10.1111/mec.15396

Brennan, P. L., Prum, R. O., McCracken, K. G., Sorenson, M. D., Wilson, R. E., & Birkhead, T. R. (2007). Coevolution of male and female genital morphology in waterfowl. *PLoS One*, *2*, e418. https://doi.org/10.1371/journal.pone.0000418

Brown, J. I., Lavretsky, P., Cumming, G., & Peters, J. L. (2019). Strong population structure and limited gene flow between Yellow-billed Ducks and mallards in southern Africa. *The Condor: Ornithological Applications*, *121*, 1–10. https://doi.org/10.1093/condor/duz042

Campagna, L., Repenning, M., Silveira, L. F., Fontana, C. S., Tubaro, P. L., & Lovette, I. J. (2017). Repeated divergent selection on pigmentation genes in a rapid finch radiation. *Science Advances*, *3*, e1602404. https://doi.org/10.1126/sciadv.1602404

Charlesworth, B. (2001). The effect of life-history and mode of inheritance on neutral genetic variability. *Genetical Research*, *77*, 153–166. https://doi.org/10.1017/s0016672301004979

DaCosta, J. M., & Sorenson, M. D. (2014). Amplification biases and consistent recovery of loci in a double-digest RAD-seq protocol. *PLoS One*, *9*, e106713. https://doi.org/10.1371/journal.pone.0106713

Dhami, K. K., Joseph, L., Roshier, D. A., & Peters, J. L. (2016). Recent speciation and elevated Z-chromosome differentiation between sexually monochromatic and dichromatic species of Australian teals. *Journal of Avian Biology*, *47*, 92–102. https://doi.org/10.1111/jav.00693

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*, 1792–1797. https://doi.org/10.1093/nar/gkh340

Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., Künstner, A., Mäkinen, H., Nadachowska-Brzyska, K., Qvarnström, A., Uebbing, S., & Wolf, J. B. W. (2012). The genomic landscape of species divergence in Ficedula flycatchers. *Nature*, *491*, 756–760. https://doi.org/10.1038/nature11584

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, *14*, 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, *164*, 1567–1587. https://doi.org/10.1093/genetics/164.4.1567

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, *180*, 977–993. https://doi.org/10.1534/genetics.108.092221

Ford, R. J., Selman, W., & Taylor, S. S. (2017). Hybridization between Mottled Ducks (*Anas fulvigula maculosa*) and mallards (*A. platyrhynchos*) in the western Gulf Coast region. *The Condor: Ornithological Applications*, *119*, 683–696. https://doi.org/10.1650/CONDOR-17-18.1

Guo, X., He, X. X., Chen, H., Wang, Z. C., Li, H. F., Wang, J. X., & Jiang, R. S. (2021). Revisiting the evolutionary history of domestic and wild ducks based on genomic analyses. *Zoological Research*, *42*, 43–50. https://doi.org/10.24272/j.issn.2095-8137.2020.133

Hammer, Ř., Harper, D. A. T., & Ryan, P. D. (2001). PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*, *4*, 1–9.

Huang, H., & Rabosky, D. L. (2015). Sex-linked genomic variation and its relationship to avian plumage dichromatism and sexual selection. *BMC Evolutionary Biology*, *15*, 99. https://doi.org/10.1186/s12862-015-0480-4

Huang, Y., Li, Y., Burt, D. W., Chen, H., Zhang, Y., Qian, W., Kim, H., Gan, S., Zhao, Y., Li, J., Yi, K., Feng, H., Zhu, P., Li, B. O., Liu, Q., Fairley, S., Magor, K. E., Du, Z., Hu, X., … Li, N. (2013). The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. *Nature Genetics*, *45*, 776–783. https://doi.org/10.1038/ng.2657

Irwin, D. E. (2018). Sex chromosomes and speciation in birds and other ZW systems. *Molecular Ecology*, *27*, 3831–3851. https://doi.org/10.1111/mec.14537

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*, 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Joseph, L., Adcock, G. J., Linde, C., Omland, K. E., Heinsohn, R., Chesser, R. T., & Roshier, D. (2009). A tangled tale of two teal: population history of the grey *Anas gracilis* and chestnut teal *A. castanea* of Australia. *Journal of Avian Biology*, *40*, 430–439. https://doi.org/10.1111/j.1600-048X.2008.04652.x

Kim, K.-W., Jackson, B. C., Zhang, H., Toews, D. P. L., Taylor, S. A., Greig, E. I., Lovette, I. J., Liu, M. M., Davison, A., Griffith, S. C., Zeng, K., & Burke, T. (2019). Genetics and evidence for balancing selection of a sex-linked colour polymorphism in a songbird. *Nature Communications*, *10*, 1852. https://doi.org/10.1038/s41467-019-09806-6

Kopelman, N. M., Mayze, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, *15*, 1179–1191. https://doi.org/10.1111/1755-0998.12387

Kraus, R. H., van Hooft, P., Megens, H. J., Tsvey, A., Fokin, S. Y., Ydenberg, R. C., & Prins, H. H. (2013). Global lack of flyway structure in a cosmopolitan bird revealed by a genome wide survey of single nucleotide polymorphisms. *Molecular Ecology*, *22*, 41–55. https://doi.org/10.1111/mec.12098

Kraus, R. H., Zeddeman, A., van Hooft, P., Sartakov, D., Soloviev, S. A., Ydenberg, R. C., & Prins, H. H. (2011). Evolution and connectivity in the world-wide migration system of the mallard: inferences from mitochondrial DNA. *BMC Genetics*, *12*, 99. https://doi.org/10.1186/1471-2156-12-99

Kreisinger, J., Munclinger, P., Javůrková, V., & Albrecht, T. (2010). Analysis of extra-pair paternity and conspecific brood parasitism in mallards *Anas platyrhynchos* using non-invasive techniques. *Journal of Avian Biology*, *41*, 551–557. https://doi.org/10.1111/J.1600-048X.2010.05002.X

Kulikova, I. V., Drovetski, S. V., Gibson, D. D., Harrigan, R. J., Rohwer, S., Sorenson, M. D., Winker, K., Zhuravlev, Y. N., & McCracken, K. G. (2005). Phylogeography of the Mallard (Anas Platyrhynchos): Hybridization, Dispersal, and Lineage Sorting Contribute to Complex Geographic Structure. *The Auk*, *122*, 949–965. https://doi.org/10.1093/auk/122.4.1309

Kulikova, I. V., Poysa, H., & Zhuravlev, Y. N. (2012). Phylogeography of the mallard *Anas platyrhynchos* from Eurasia inferred from sequencing of the mtDNA control region. *Russian Journal of Genetics*, *48*, 705–712. https://doi.org/10.1134/S1022795412070095

Kulikova, I. V., & Zhuravlev, Y. N. (2009). Molecular genetic studies of Anatinae interspecific hybridization on the example of *Anas platyrhynchos* supraspecific complex. *Advances in Contemporary Biology*, *129*, 158–167. (in Russian)

Kulikova, I. V., Zhuravlev, Y. N., & McCracken, K. G. (2004). Asymmetric hybridization and sex-biased gene flow between Eastern spot-billed ducks (*Anas zonorhyncha*) and mallards (*A. platyrhynchos*) in the Russian Far East. *The Auk*, *121*, 930–949. https://doi.org/10.1093/auk/121.3.930

Lamichhaney, S., Han, F., Webster, M. T., Grant, B. R., Grant, P. R., & Andersson, L. (2020). Female-biased gene flow between two species of Darwin's finches. *Nature Ecology and Evolution*, *4*, 979–986. https://doi.org/10.1038/s41559-020-1183-9

Lavretsky, P., DaCosta, J. M., Hernández-Baños, B. E., Engilis, A. Jr, Sorenson, M. D., & Peters, J. L. (2015). Speciation genomics and a role for the Z chromosome in the early stages of divergence between Mexican ducks and mallards. *Molecular Ecology*, *24*, 5364–5378. https://doi.org/10.1111/mec.13402

Lavretsky, P., DaCosta, J. M., Sorenson, M. D., McCracken, K. G., & Peters, J. L. (2019). ddRAD-seq data reveal significant genome-wide population structure and divergent genomic regions that distinguish the mallard and close relatives in North America. *Molecular Ecology*, *28*, 2594–2609. https://doi.org/10.1111/mec.15091

Lavretsky, P., Engilis, A. Jr, Eadie, J. M., & Peters, J. L. (2015). Genetic admixture supports an ancient hybrid origin of the endangered Hawaiian duck. *Journal of Evolutionary Biology*, *28*, 1005–1015. https://doi.org/10.1111/jeb.12637

Lavretsky, P., Hernández-Baños, B. E., & Peters, J. L. (2014). Rapid radiation and hybridization contribute to weak differentiation and hinder phylogenetic inferences in the New World mallard complex (*Anas* spp.). *The Auk: Ornithological Advances*, *131*, 524–538. https://doi.org/10.1642/AUK-13-164.1

Lavretsky, P., Janzen, T., & McCracken, K. G. (2019). Identifying hybrids & the genomics of hybridization: Mallards & American black ducks of Eastern North America. *Ecology and Evolution*, *9*, 3470–3490. https://doi.org/10.1002/ece3.4981

Lavretsky, P., McCracken, K. G., & Peters, J. L. (2014). Phylogenetics of a recent radiation in the mallards and allies (Aves: Anas): Inferences from a genomic transect and the multispecies coalescent. *Molecular Phylogenetics and Evolution*, *70*, 402–411. https://doi.org/10.1016/j.ympev.2013.08.008

Lavretsky, P., McInerney, N. R., Mohl, J. E., Brown, J. I., James, H. F., McCracken, K. G., & Fleischer, R. C. (2020). Assessing changes in genomic divergence following a century of human-mediated secondary contact among wild and captive-bred ducks. *Molecular Ecology*, *29*, 578–595. https://doi.org/10.1111/mec.15343

Lima, T. G. (2014). Higher levels of sex chromosome heteromorphism are associated with markedly stronger reproductive isolation. *Nature Communications*, *5*, 4743. https://doi.org/10.1038/ncomms5743

Mank, J. E., Nam, K., & Ellegren, H. (2010). Faster-Z evolution is predominantly due to genetic drift. *Molecular Biology and Evolution*, *27*, 661–670. https://doi.org/10.1093/molbev/msp282

Mayr, E. (1982). Of what use are subspecies? *The Auk*, *99*, 593–595.

McCracken, K. G., Johnson, W. P., & Sheldon, F. H. (2001). Molecular population genetics, phylogeography, and conservation biology of the Mottled Duck (*Anas fulvigula*). *Conservation Genetics*, *2*, 87–102. https://doi.org/10.1023/A:1011858312115

McKinney, F., Derrickson, S. R., & Mineau, P. (1983). Forced copulation in waterfowl. *Behaviour*, *86*, 250–294. https://doi.org/10.1163/156853983X00390

Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America*, *76*, 5269–5273. https://doi.org/10.1073/pnas.76.10.5269

Noor, M. A., & Feder, J. L. (2006). Speciation genetics: Evolving approaches. *Nature Reviews Genetics*, *7*, 851–861. https://doi.org/10.1038/nrg1968

Peters, J. L., Lavretsky, P., DaCosta, J. M., Bielefeld, R. R., Feddersen, J. C., & Sorenson, M. D. (2016). Population genomic data delineate conservation units in mottled ducks (*Anas fulvigula*). *Biological Conservation*, *203*, 272–281. https://doi.org/10.1016/j.biocon.2016.10.003

Peters, J., Sonsthagen, S. A., Lavretsky, P., Rezsutek, M. J., Johnson, W. P., & McCracken, K. G. (2013). Interspecific hybridization contributes to high genetic diversity and apparent effective population size in an endemic population of mottled ducks (*Anas fulvigula maculosa*). *Conservation Genetics*, *15*, 509–520. https://doi.org/10.1007/s10592-013-0557-9

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: And inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*, *7*, e37135. https://doi.org/10.1371/journal.pone.0037135

Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., & Lercher, M. J. (2014). PopGenome: An efficient swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution*, *31*, 1929–1936. https://doi.org/10.1093/molbev/msu136

Polivanova, N. N. (1971). Birds of Lake Khanka, part 1 (game waterfowl and colonial birds). In V. M. Polivanov (Ed.), *Proceedings of the Kedrovaya Pad Nature Reserve*, Vol. *3*. Vladivostok, Russia: Academy of Science Publishing [In Russian].

Price, J. J. (2019). Sex differences in song and plumage color do not evolve through sexual selection alone: New insights from recent research. *Journal of Ornithology*, *160*, 1213–1219. https://doi.org/10.1007/s10336-019-01681-8

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*, 559–575. https://doi.org/10.1086/519795

Rhymer, J. M., Williams, M. J., & Braun, M. J. (1994). Mitochondrial analysis of gene flow between New Zealand mallards (*Anas platyrhynchos*) and Grey Ducks (*A. superciliosa*). *The Auk*, *111*, 970–978. https://doi.org/10.2307/4088829

Skryabin, N. G. (1963). Spot-billed duck range expansion in Pribaikalye. In V. F. Larionov (Ed.), *Ornithology* (pp. 311–314). Moscow State University Press. [In Russian]

Söderquist, P., Elmberg, J., Gunnarsson, G., Thulin, C.-G., Champagnon, J., Guillemain, M., Kreisinger, J., Prins, H. H. T., Crooijmans, R. P. M. A., & Kraus, R. H. S. (2017). Admixture between released and wild game birds: A changing genetic landscape in European mallards (*Anas platyrhynchos*). *European Journal of Wildlife Research*, *63*, 98. https://doi.org/10.1007/s10344-017-1156-8

Toews, D. P., Taylor, S. A., Vallender, R., Brelsford, A., Butcher, B. G., Messer, P. W., & Lovette, I. J. (2016). Plumage genes and little else distinguish the genomes of hybridizing warblers. *Current Biology*, *26*, 2313–2318. https://doi.org/10.1016/j.cub.2016.06.034

Toomey, M. B., Marques, C., Andrade, P., Araújo, P. M., Sabatino, S. J., Gazda, M. A., & Carneiro, M. (2018). A non-coding region near Follistatin controls head colour polymorphism in the Gouldian finch. *Proceedings of the Royal Society B*, *285*(1888), 20181788. https://doi.org/10.1098/rspb.2018.1788

Wang, W., Wang, Y., Lei, F., Liu, Y., Wang, H., & Chen, J. (2019). Incomplete lineage sorting and introgression in the diversification of Chinese spot-billed ducks and mallards. *Current Zoology*, *65*, 589–597. https://doi.org/10.1093/cz/zoy074

Wells, C. P., Lavretsky, P., Sorenson, M. D., Peters, J. L., DaCosta, J. M., Turnbull, S., Uyehara, K. J., Malachowski, C. P., Dugger, B. D., Eadie, J. M., & Engilis, A. (2019). Persistence of an endangered native duck, feral mallards, and multiple hybrid swarms across the main Hawaiian Islands. *Molecular Ecology*, *28*, 5203–5216. https://doi.org/10.1111/mec.15286

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.

Wright, A. E., Harrison, P. W., Zimmer, F., Montgomery, S. H., Pointer, M. A., & Mank, J. E. (2015). Variation in promiscuity and sexual selection drives avian rate of Faster-Z evolution. *Molecular Ecology*, *24*, 1218–1235. https://doi.org/10.1111/mec.13113

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Kulikova, I. V., Shedko, S. V., Zhuravlev, Y. N., Lavretsky, P. & Peters, J. L. (2022). Z-chromosome outliers as diagnostic markers to discriminate Mallard and Chinese Spot-billed Duck (Anatidae). *Zoologica Scripta*, *00*, 1–14. https://doi.org/10.1111/zsc.12540